The Importance of Standard Image Formats for Scientific Progress

Nicholas K. Sauter, Physical Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720

Standardization of the crystallographic image format, as embodied by the imgCIF library, can have a positive role supporting the development of new detector technology, and making macromolecular diffraction data more accessible to the scientific community. The diversity of image file formats produced by different vendors poses a variety of problems for writing data processing programs such as *LABELIT* (1). A standard format, if it is flexible enough to support all anticipated experimental configurations, offers the advantage that it centralizes the software needed to decode the image. Having previously implemented an imgCIF wrapper for *LABELIT*, it is now easy to extend support to new detectors that use imgCIF as a native format. Thus the new Pilatus 6M (2), while it represents a new detector technology and an unusual pixel layout, can be plugged in to an existing data processing pipeline. Interest in creating permanent data archives has recently been renewed (3), suggesting an obvious role for centralized software that allows data to be retrieved reliably in the future. Entire raw datasets often contain information that is not captured in the processed structure factor file deposited with the Protein Data Bank. For example, if multiple crystal samples are studied, symmetry elements may be non-crystallographic in one sample, but exact in another sample. Certain datasets reveal non-merohedral twinning under close examination. Taking such factors into account in crystallographic structure refinement may lead to an improved understanding of particular structures.

(1) Sauter, N. K., Grosse-Kunstleve, R. W. & Adams, P. D. (2004). *J. Appl. Cryst.* **37**, 399-409.

(2) Hülsen, G., Broennimann, C., Eikenberry, E. F. & Wagner, A. (2006) *J. Appl. Cryst.* **39**, 550-557.

(3) Baker, E., Dauter, Z., Guss, M. & Einspahr, H. (2008). *Acta Cryst.* D**64**, 337-338.